公募研究：2002～2003年度

転写因子のターゲット部位予測にもとづく遺伝子の機能及び相互作用推定

●河野　秀俊[1,2)]　◆Kim Thi Phuong Oanh[2)]

日本原子力研究所　1)中性子利用研究センター　2）計算科学技術推進センター

〈研究の目的と進め方〉

ゲノム科学は、ポストシークエンス、すなわち蓄積された配列情報から遺伝子機能の解析へと展開してきている。生物はさまざまな遺伝子の相互作用ネットワークの上に成り立っており、機能の解析にはそのネットワークを調べる必要がある。遺伝子の発現は、複数の転写因子によって制御されている。従って、ゲノムの機能を推定するには、どの転写因子がどの遺伝子の発現を制御しているか調べる必要がある。本研究では、１）計算科学的なアプローチから、転写因子それぞれのゲノム上でのターゲット部位を予測する、２）共通の転写因子をもつ遺伝子をクラスタリングすることによって、遺伝子間の相互作用を推定し、実験結果と比較検証し、仮定･予測方法の妥当性を評価し、遺伝子間の関連付けを行なうことを目指す。

〈研究開始時の研究計画〉

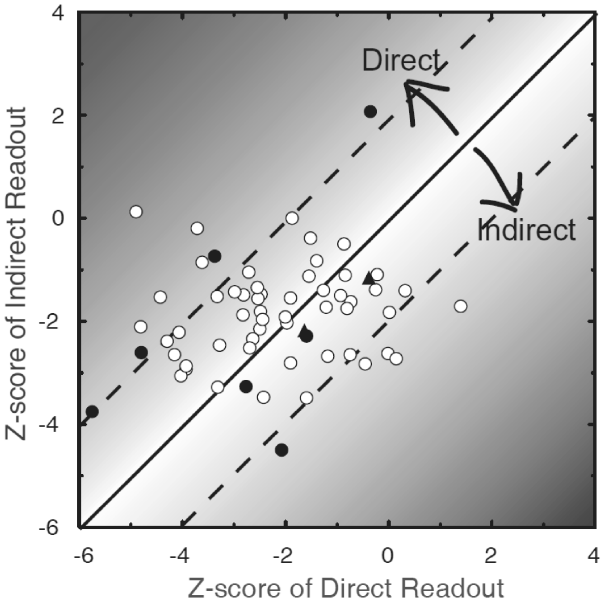
1. 蛋白質 - ＤＮＡ複合体構造から塩基 - アミノ酸相互作用のコンタクトポテンシャル（直接認識）と複合体構造に対するＤＮＡ配列適合度を測るポテンシャル（間接認識）を作成する。そのために、ＰＤＢから複合体構造を自動的に導出し定期的にアップデートするシステムを構築する。
2. 更新された二つのポテンシャルを組み合わせることによって、ターゲット配列とそれ以外の配列の判別能力を向上させる。
3. 蛋白質のフォールドタイプを探すスレッディング(いわゆる３Ｄ-１Ｄ法)のアナロジーにより、転写因子のターゲット部位をそれぞれの遺伝子の上流領域から探す。そして、下流域にある遺伝子と転写因子の対応関係を調べる。
4. 遺伝子の機能、相互作用の推定は、「同じ転写因子によって制御されている遺伝子は、機能的に関係が深く、共同的に一連のパスウェイ上で働くことが多い」という経験則を検証する。転写制御についてよく研究されている酵母のゲノムに対して遺伝子の機能及び相互作用の推定を行う。機能既知の遺伝子に対する推定結果を評価することにより、この経験則の妥当性を検証する。
5. さらに他の生物種のゲノムでも同様な計算を行うことにより、ある生物種で推定された遺伝子産物と相同な蛋白質を探すことにより、機能及び相互作用の推定ができる遺伝子の数を増やす。

〈研究期間の成果〉

1）年々増加する蛋白質-ＤＮＡ複合体の立体構造に対して､自動的にデータを更新できるシステムを構築した。これにより、167の配列冗長性のない蛋白質-ＤＮＡの複合体構造をＰＤＢから抽出し、直接認識、間接認識ポテンシャルを更新した。

2）ＤＮＡ結合蛋白質は、ゲノムのＯＲＦの上流・下流に結合して、転写を制御していることが知られている。そこで、任意の長さの上流・下流のＤＮＡ配列を抽出することを行った。その抽出した配列に対して、更新したポテンシャルを用いて、蛋白質のフォールドタイプを探すスレッディング法のアナロジーにより、ＤＮＡ結合蛋白質のターゲット部位を探せるようにプログラムを整備した。

3）ポテンシャルを用いてさまざまなＤＮＡ結合蛋白質のＤＮＡ結合様式を直接、間接認識の面から評価した。その結果、図に示すように、その寄与バランスはＤＮＡ結合蛋白質依存であり、構造モチーフやファミリ内であきらかな関係は見つからなかった。制限酵素のような非常に配列特異性の強い蛋白質でも、直接認識と間接認識の寄与は蛋白質種依存であることがわかった。また、間接認識に比べて直接認識のZ-scoreは蛋白質間の分散が大きいことがわかった[4]。これは、蛋白質によって読み取られるターゲット配列の物理化学的情報は、配列が異なるにも関わらずほぼ同程度であることを示唆している。



また、個々の蛋白質-ＤＮＡ複合体において、直接認識エネルギーと間接認識エネルギーには相関は見られなかった。このことから、両者は独立した情報をもっているものと考えた。実際、両エネルギーを足し合わせることにより、ターゲット部位の予測精度を上げることができた[4,5]。また、複数の転写因子がＤＮＡに結合する場合、単独でＤＮＡに結合する場合よりも配列選択性が強くなり、協同的にターゲット配列を認識していることを定量的に示した[4,5,6]。

4）実用レベルでの有用性を検証するために､ゲノム上でのターゲット配列を探した。直接認識と間接認識ポテンシャルを組み合わせて、酵母のゲノムに対して、ＭＡＴα2/ＭＣＭ１ヘテロダイマー蛋白質のターゲット部位を予測した。ＤＮＡ結合蛋白質は、ゲノムのＯ

ＲＦの上流・下流に結合して、転写を制御していることが知られている。そこで、酵母のすべての遺伝子の上流１kbp長のＤＮＡ配列を抽出した。その抽出した配列に対して、更新したポテンシャルを用いて、蛋白質のフォールドタイプを探すスレッディング法のアナロジーにより、ＤＮＡ結合蛋白質のターゲット部位を探した。その結果、ＭＡＴα２/ＭＣＭ１のヘテロダイマー蛋白質が結合すると予測された部位のうち、Z-scoreの上位6つは実験的にも結合が確かめられた部位であった。単純なモチーフサーチでは結合部位と予測されるものの、実験的にターゲット部位でないことが明らかにされている部位に対し、本方法は結合可能性が低いと予測し、実験結果を正しく再現していた。単純な配列のモチーフサーチでは、フォルスポジティブが多く生じる。本方法では、モチーフのパターンを定量化し、その前後の配列特性も考慮することにより、フォルスポジティブを抑えた予測をすることができた。予測されたZ-scoreの上位６つの部位の下流にある遺伝子はすべて酵母の性決定に関わるα因子もしくはα因子に直接的または間接的に関与するものであった。このケースでは、遺伝子の機能とそれを制御している蛋白質に明らかな関係が見られた[3]。

〈国内外での成果の位置づけ〉

これまでの転写因子のターゲット部位予測は、TRANSFACデータベースに集積された配列に基づく方法がほとんどで、本研究のようなに転写因子と複合体の立体構造から転写因子のターゲットＤＮＡ配列を予測する汎用的なものはなかった。

Klugら(1999) は、立体構造から導出した単純なルールでZnフィンガーモチーフを持つ転写因子のターゲット配列の予測に成功しているが、これはＤＮＡ結合蛋白質一般への汎用性がない。近年、申請者らと同様に、ＤＮＡ蛋白質の結合配列特異性を、アミノ酸－ＤＮＡの直接的な相互作用とＤＮＡ構造の配列依存性の面から計算した論文がLaveryら(Structure, 2004)によって報告されているが、ゲノムワイドに適用できるような方法ではない。

申請者らが開発した転写因子のターゲット部位を推定する法は、蛋白質‐ＤＮＡ複合体の立体構造から得られるアミノ酸と塩基の相互作用及びＤＮＡ構造の配列依存性にもとづいており、立体構造がモデリングできる蛋白質すべてに適用できる汎用的な方法である[3]。また、立体構造が得られれば（これは、往々にして未だに難しいが)、ＤＮＡ結合蛋白質のターゲット配列を探す膨大でコストのかかる実験をすることなく、構造をもとにターゲット配列を推定することができる。申請者らの成果は、構造生物学のバイオインフォマティックスとして高い評価を受け、2005年に2つのレビュー誌、Annu. Rev. Biophys. Biomol. Str. [2]とGene [1]から招待論文の形で発表された。

〈達成できなかったこと、予想外の困難、その理由〉

蛋白質‐ＤＮＡ複合体の立体構造から、統計ポテンシャルを自動的に導出し、定期的に更新するシステムを構築した。しかし、蛋白質とＤＮＡの結合が特定的か非特異的かを判断するのは容易でなく、ひとつひとつ文献でチェックする必要があり、この作業に非常に時間がかかった。また、同じ転写因子で制御される遺伝子を推定しても、それが正しいかどうかひとつずつ検討するのにデータベースや文献を丁寧にあたる必要があり、予想以上に推定結果の検証に時間を要した。各段階で研究者による文献チェックが必要なため、ポテンシャル作成の完全な自動化はできなかった。

ＤＮＡ配列そのものの配列情報(直接情報)とＤＮＡがもつ物理化学的な特性である間接情報を読み取るために、それぞれ直接認識、間接認識ポテンシャルを構築し、それらを組み合わせることにより、より精度のよい転写因子のターゲット部位予測を実現することができた。ターゲット部位の予測結果にもとづいた遺伝子の機能推定は、大部分の場合、関連付けを行なっただけにとどまり、個々の関連付けに対して詳細な検討をすることができなかった。これは、膨大な関連付け情報に対して、実験的なデータとの対応を自動的につけることができなかったためである。

〈今後の課題〉

ターゲット部位の予測結果にもとづいた遺伝子の機能推定は、大部分の場合、関連付けを行なっただけにとどまり、個々の関連付けに対して詳細な検討をすることができなかった。このことを解決するために、今後はこの対応付けをいかに効率的かつ正確に行なうかが遺伝子間の機能統合を行なう上での課題になると思われる。ひとつの解決策として、オントロジー研究の今後の発展に期待する。

ある生物種に存在するすべてのＤＮＡ結合蛋白質の結合部位を予測することで、その生物のもつ転写制御関係が明らかにされる。できるだけ多くのＤＮＡ結合蛋白質のターゲット配列を推定するために、複合体の立体構造が解かれていない蛋白質に関して、正確かつ迅速に複合体のモデルを作ることが今後の課題である。それができれば、それぞれの蛋白質によって制御されている遺伝子推定が実行でき、遺伝子間の関係が明らかにできるものと考えている。また、ＤＮＡ結合蛋白質とＤＮＡの相互作用のみならず、蛋白質-蛋白質相互作用による転写因子同士の相互作用も考慮することで、遺伝子制御ネットワークをより正確に詳細に明らかにしていけると考える。

生物種間の転写因子と遺伝子の制御がわかってくると、転写因子自体の進化に対しても、踏み込んだ研究が行えると考える。つまり、制御システムがどのように生み出され、進化してきたかということを研究することは、生命現象の理解だけでなく、新しい人工システムの創造において極めて重要であると考える。

〈研究期間の全成果公表リスト〉

1) 論文

1. Gromiha, M. M., Siebers, J .G., Selvaraj, S., Kono H. and Sarai A., Role of Inter and Intramolecular Interactions in Protein-DNA Recognition, Gene, 364, 108-113 (2005)
2. Sarai, A. and Kono, H., Protein-DNA Recognition Patterns and Predictions, Annu. Rev. Biophys. Biomol. Struct. 34, 379-398 (2005)
3. Sarai, A., Siebers, J. G., Selvaraj, S., Gromiha, M.M. and Kono, H., Integration of bioinformatics and computational biology to understand protein-DNA recognition mechanism, J. Bioinform. Comput. Biol. 3, 1-15 (2005)
4. 404010058
Gromiha, M. M., Siebers, J. G., Selvaraj, S., Kono, H. and Sarai, A. Intermolecular and Intramolecular Readout

Mechanisms in Protein-DNA Recognition J. Mol. Biol. 337, 285-294 (2004)
5. 404010111
Sarai, A., Selvaraj, S., Gromiha, M. M., and Kono, H. Structure-Function Relationship in DNA Sequence Recognition by Transcription Factors, Bioinformatics 2004, Second Asia Pacific Bioinformatics Conference (APBC2004) 233-238 (2004)
6. 404010118
Kono, H. and Sarai, A. Structure-Specificity Relationship in Protein-DNA Recognition, Bioinformatics of Genome Regulation and Structure eds. Kolchanov and Hofestaedt, R., Kluwer 155-161 (2004)